CLAIMS

What is claimed is:

1. A library of expression vectors encoding a library of protein complexes, each vector comprising:

a first nucleotide sequence encoding a first polypeptide subunit; and

a second nucleotide sequence encoding a second polypeptide subunit; wherein

the first and second nucleotide sequences each independently varies within the library of expression vectors, and

the first polypeptide subunit and the second polypeptide subunit are expressed as separate proteins which self-assemble to form a protein complex in cells into which the library of expression vectors are introduced.

2. The library of claim 1, wherein the expression vector is a yeast expression vector.

3. The library of claim 1, wherein the expression vector is a yeast-bacterial shuttle vector which contains a bacterial origin of replication.

4. The library of claim 1, wherein the expression vector is a bacterial plasmid.

5. The library of claim 1, wherein the expression vector is a mammalian expression vector.

6. The library of claim 1, wherein the expression vector is a viral vector selected from the group consisting of adenovirus, adeno-associated virus,

vaccinia, retrovirus, and herpes simplex virus vectors.

7. The library of claim 1, wherein the diversity of the first and the second polypeptide subunits each independently is at least 10^3.

8. The library of claim 1, wherein the diversity of the first or the second polypeptide subunits each independently is at least 10^5.

9. The library of claim 1, wherein the diversity of the protein complexes encoded by the library of expression vectors is at least $1x10^7$.

10. The library of claim 1, wherein the diversity of the protein complexes encoded by the library of expression vectors is at least $1x10^{10}$.

11. The library of claim 1, wherein the diversity of the protein complexes encoded by the library of expression vectors is at least $1x10^{12}$.

12. The library of claim 1, wherein the diversities of the first and second polypeptide subunits are each independently derived from libraries of precursor sequences that are not specifically designed for a particular target peptide or protein.

13. The library of claim 1, wherein the diversities of the first and second polypeptide subunits are not derived from one or more proteins that are known to bind to a particular target peptide or protein.

14. The library of claim 1, wherein the diversities of the first and second polypeptide subunits are not generated by mutagenizing one or more proteins that are known to bind to a particular target peptide or protein.

15. The library of claim 1, wherein the first and the second polypeptide subunits are subunits of a multimeric protein whose sequence varies within a library of multimeric proteins.

16. The library of claim 15, wherein the library of multimeric proteins are selected from the group consisting of libraries of antibodies, growth factor receptors, T cell receptors, cytokine receptors, tyrosine kinase-associated receptors, and MHC proteins.

17. The library of claim 1, wherein the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody heavy-chain region, and the second nucleotide sequence comprises a coding sequence of an antibody light-chain region.

18. The library of claim 1, wherein the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody heavy-chain variable region, and the second nucleotide sequence comprises a coding sequence of an antibody light-chain variable region.

19. The library of claim 1, wherein the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody heavy-chain variable and constant 1 region, and the second nucleotide sequence comprises a coding sequence of an antibody light-chain variable and constant region.

20. The library of claim 17, wherein the source of the coding sequences of the antibody light-chain and heavy-chain regions is from human, non-human primate, or rodent DNA.

21. The library of claim 17, wherein the source of the coding sequences of the antibody light-chain and heavy-chain variable regions is from one or more non-immunized animals.

22. The library of claim 17, wherein the source of the coding sequences of the antibody light-chain and heavy-chain variable regions are selected from the group consisting of human fetal spleen, fetal liver, bone marrow, lymph nodes and peripheral blood cells.

23. The library of claim 1, wherein the first and second polypeptide subunits each further comprises a plurality of cysteine residues adjacent the N- or C- terminus.

24. The library of claim 23, wherein the plurality of cysteine residues are 2-8 cysteine residues.

25. The library of claim 1, wherein the first and second polypeptide subunits each further comprises a first and second zipper domain, respectively, which brings first and second polypeptide subunits into close proximity through non-covalent interactions between the first and second zipper domain.

26. The library of claim 25, wherein the first and second zipper domain each is a leucine zipper.

27. The library of claim 25, wherein the leucine zipper is selected from the group consisting of the Myc, Max, Jun, Fos, and neural cadherin leucine zippers.

28. The library of claim 1, wherein the first or second polypeptide subunit further comprises a bundle domain which brings into close proximity a plurality of protein complexes each of which is formed between the first and second polypeptide subunits through non-covalent interactions between the bundle domains.

29. The library of claim 28, wherein the bundle domain is a coiled-coil assembly domain of a protein selected from the group consisting of cartilage oligomeric matrix protein, the NUDE protein, and bacteriphage T4 fribritin.

30. The library of claim 28, wherein the bundle domain is fused to the C-terminus of the first or second polypeptide subunit.

31. The library of claim 28, wherein the bundle domain is linked to the C-terminus of the first or second polypeptide subunit through a peptide linker.

32. The library of claim 30, wherein the peptide linker comprises SEQ ID NO: 79.

33. The library of claim 1, wherein each of the expression vectors further comprises a sequence encoding an affinity tag fused with the first or second polypeptide subunit.

34. The library of claim 33, wherein the affinity tag is selected from the group consisting of a polyhistidine tag, polyarginine tag, glutathione-S-transferase , maltose binding protein, staphylococcal protein A tag, and an EE-epitope tag.

35. The library of claim 1, wherein the first polypeptide subunit further comprises a transcription sequence encoding an activation domain or a DNA binding domain of a transcription activator.

36. The library of claim 35, wherein the transcription sequence is 5' relative to the first nucleotide sequence.

37. The library of claim 35, wherein the transcription sequence is 3' relative to the first nucleotide sequence.

38. The library of claim 35, wherein the transcription activator is a transcription activator having separable DNA-binding and transcription activation domains.

39. The library of claim 35, wherein the transcription activator is selected from the group consisting of GAL4, GCN4, and ADR1 transcription activator.

40. The library of claim 1, wherein the first or second polypeptide subunit further comprises a Ras guanyl nucleotide exchange factor (SOS factor).

41. The library of claim 1, wherein the first or second polypeptide subunit further comprises a membrane targeting signal.

42. The library of claim 41, wherein the membrane targeting signal is a myristoylation sequence.

43. The library of claim 41, wherein the membrane targeting signal is a farnesylation sequence.

44. The library of claim 1, wherein the first or second polypeptide subunit further comprises a portion of mammalian Ras lacking the carboxy-terminal domain (the CAAX box).

45. The library of claim 1, wherein the first or second polypeptide subunit further comprises a ubiquitin sequence.

46. The library of claim 1, where the first or second polypeptide subunit further comprises a yeast agglutinin cell wall protein which facilitates transportation of the library protein complexes to the surface of yeast cells.

47. The library of claim 1, where the yeast agglutinin cell wall protein is Aga2p yeast cell wall protein.

48. The library of claim 1, wherein expression of the first and second polypeptide subunits is controlled by separate promoters.

49. The library of claim 1, wherein the first and second polypeptide subunits are expressed bicistronically from the same promoter.

50. A library of transformed yeast cells, comprising: yeast cells transformed a library of yeast expression vectors, each vector comprising

a first nucleotide sequence encoding a first polypeptide subunit; and

a second nucleotide sequence encoding a second polypeptide subunit; wherein

the first and second nucleotide sequences each independently varies within the library of expression vectors, and

the first polypeptide subunit and the second polypeptide subunit are expressed as separate proteins which self aggregate to form a protein

complex in cells into which the library of expression vectors are introduced.

51. The library of claim 50, wherein the yeast cells are diploid yeast cells.

52. The library of claim 50, wherein the yeast cells are haploid yeast cells.

53. The library of claim 52, wherein the haploid yeast cells are of a or α strain of yeast.

54. A method for generating a library of yeast expression vectors, comprising:

transforming into yeast cells a library of insert nucleotide sequences that are linear and double-stranded, and a library of linearized yeast expression vectors, each having a 5'- and 3'- terminus sequence at the site of linearization; and

having homologous recombination occur between the vector and the insert sequence such that the insert sequence is included in the vector in the transformed yeast cells,

wherein

each of the linearized yeast expression vectors in the vector library comprises a first polynucleotide sequence encoding a first polypeptide subunit which varies within the vector library;

the insert sequences of the insert library comprise a second nucleotide sequence encoding a second polypeptide subunit which varies within the insert library, each of the insert sequences comprising a 5'- and 3'- flanking sequence at the respective ends of the insert sequence and being sufficiently homologous to the 5'- and 3'-terminus sequences of the linearized yeast expression vector, respectively, to enable homologous recombination to occur, and

the first and second polypeptide subunits are expressed as separate proteins.

55. The method of claim 54, wherein the 5'- or 3'- flanking sequence of the insert nucleotide sequence is between about 20-120 bp in length.

56. The method of claim 54, wherein the 5'- or 3'- flanking sequence of the insert nucleotide sequence is between about 40-90 bp in length.

57. The method of claim 54, wherein the 5'- or 3'- flanking sequence of the insert nucleotide sequence is between about 45-55 bp in length.

58. The method of claim 54, wherein the yeast expression vector is a 2μ plasmid vector.

59. The method of claim 54, wherein the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody heavy chain region, and the second nucleotide sequence comprises a coding sequence of an antibody light chain region.

60. The method of claim 54, wherein the library of insert nucleotide sequences are inserted into a site of the vector such that expression of the first and second polypeptide subunits is under the transcriptional control of separate promoters.

61. The method of claim 54, wherein the library of insert nucleotide sequences are inserted into a site of the vector such that the first and second polypeptide subunits are expressed bicistronically from the same promoter.

61. The method of claim 54, wherein the first and second polypeptide subunits when expressed are expressed as separate proteins that self-assemble in cells into which the library of expression vectors are introduced.

63. A method of producing a library of antibodies or antibody fragments, comprising:

expressing in yeast cells a library of yeast expression vectors, each vector comprising

a first nucleotide sequence encoding an antibody heavy chain region, and

a second nucleotide sequence encoding an antibody light chain region,

wherein

the antibody heavy chain region and the antibody light chain region each independently varies within the library of expression vectors, and

the antibody heavy chain region and the antibody light chain region are expressed as separate proteins which self-assemble in the yeast cells to form an antibody or an antibody fragment.

64. The method of claim 63, wherein the diversity of the library of antibodies or antibody fragments is between about $1x10^{7}$-$1x10^{18}$.

65. The method of claim 63, wherein the diversity of the library of antibodies or antibody fragments is between about $1x10^{8}$-$1x10^{18}$.

66. The method of claim 63, wherein the diversity of the library of antibodies or antibody fragments is between about $1x10^{12}$-$1x10^{18}$.

67. The method of claim 63, wherein the first nucleotide sequence encodes

an antibody heavy-chain variable region, and the second nucleotide sequence encodes an antibody light-chain variable region.

68. The library of claim 63, wherein the first nucleotide sequence encodes an antibody heavy-chain variable and constant 1 region, and the second nucleotide sequence encodes an antibody light-chain variable and constant region.

69. A method for selecting tester protein complexes capable of binding to a target peptide or protein, the method comprising:

expressing a library of tester protein complexes in yeast cells, each tester protein complex being formed between a first polypeptide subunit whose sequence varies within the library and a second polypeptide subunit which is expressed as a separate protein from the first polypeptide subunit and whose sequence varies within the library independently of the first polypeptide;

expressing a target fusion protein in the yeast cells expressing the tester protein complexes, the target fusion protein comprising a target peptide or protein; and

selecting those yeast cells in which a reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester protein complex to the target fusion protein.

70. The method of claim 69, wherein expressing the library of tester protein complexes includes

transforming a library of tester expression vectors into the yeast cells which contain a reporter construct comprising the reporter gene whose expression is under transcriptional control of a transcription activator comprising an activation domain and a DNA binding domain, each tester

expression vector comprising

a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator,

a first nucleotide sequence encoding the first polypeptide subunit fused which is expressed as a fusion protein with either the activation domain or the DNA binding domain of the transcription activator, and

a second nucleotide sequence encoding the second polypeptide subunit which is expressed as a separate protein from the first polypeptide subunit.

71. The method of claim 70, wherein expressing a target fusion protein includes

transforming a target expression vector into the yeast cells simultaneously or sequentially with the library of tester expression vectors, the target expression vector comprising

a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and

a target sequence encoding the target protein or peptide; and

expressing the target fusion protein from the target expression vector.

72. The method of claim 69, wherein the steps of expressing the library of tester protein complexes and expressing the target fusion protein include causing mating between first and second populations of haploid yeast cells of opposite mating types,

wherein

the first population of haploid yeast cells comprises

a library of tester expression vectors for the library of tester fusion proteins, each tester expression vector comprising

a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator,

a first nucleotide sequence encoding the first polypeptide subunit fused which is expression as a fusion protein with either the activation domain or the DNA binding domain of the transcription activator, and

a second nucleotide sequence encoding the second polypeptide subunit which is expressed as a separate protein from the first polypeptide subunit; and

the second population of haploid yeast cells comprises a target expression vector comprising

a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors, and

a target sequence encoding the target protein or peptide; and

either the first or second population of haploid yeast cells comprises a reporter construct comprising the reporter gene whose expression is under transcriptional control of the transcription activator.

73. The method of claim 72, wherein the haploid yeast cells of opposite mating types are α and a type strains of yeast.

74. The method of claim 73, wherein the mating between the first and second populations of haploid yeast cells of α and a type strains is in a rich nutritional culture medium.

75. The method of claim 69, wherein the diversity of the protein complexes encoded by the library of yeast expression vectors is at least $1x10^7$.

76. The method of claim 69, wherein the diversity of the protein complexes

encoded by the library of yeast expression vectors is at least $1x10^{10}$.

77. The method of claim 68, wherein the diversity of the protein complexes encoded by the library of yeast expression vectors is at least $1x10^{12}$.

78. The method of claim 68, wherein the first nucleotide sequence in the library of expression vectors comprises a coding sequence of an antibody light-chain region, and the second nucleotide sequence comprises a coding sequence of an antibody heavy-chain region.

79. The method of claim 68, wherein the conformation of the protein complexes expressed by the library of expression vectors mimics a conformation of an antibody.

80. The method of claim 68, further comprising:

isolating the tester expression vector from the selected clones; and

mutagenizing the first and second nucleotide sequences in the isolated tester expression vectors to form a library of mutagenized expression vectors.

81. The method of claim 80, wherein the mutagenesis is selected from the group consisting of error-prone PCR mutagenesis, site-directed mutagenesis, DNA shuffling and combinations thereof.

82. The method of claim 68, wherein the target fusion protein comprises an antigen associated with a disease state.

83. The method of claim 68, wherein the target fusion protein comprises a tumor-surface antigen.

84. The method of claim 69, wherein the target fusion protein comprises a human growth factor receptor.

85. The method of claim 84, wherein the human growth factor is selected from the group consisting of epidermal growth factors, transferrin, insulin-like growth factor, transforming growth factors, interleukin-1, and interleukin-2.

86. The method of claim 69, wherein the protein encoded by the reporter gene is selected from the group consisting of β-galactosidase, α-galactosidase, luciferase, β-glucuronidase, chloramphenicol acetyl transferase, secreted embryonic alkaline phosphatase, green fluorescent protein, enhanced blue fluorescent protein, enhanced yellow fluorescent protein, and enhanced cyan fluorescent protein.

87. A method for selecting tester proteins capable of binding to a target peptide or protein, the method comprising:

expressing a library of tester protein complexes in yeast cells, each tester protein complex being formed in vivo between a first polypeptide subunit whose sequence varies within the library and a second polypeptide subunit which is expressed as a separate protein from the first polypeptide subunit and whose sequence varies within the library independently of the first polypeptide;

expressing a plurality of target fusion proteins in the yeast cells expressing the tester proteins, each of the target fusion proteins comprising a target peptide or protein; and

selecting those yeast cells in which a reporter gene is expressed, the expression of the reporter gene being activated by binding of the tester fusion to the target fusion protein.

88. The method of claim 87, wherein the steps of expressing the library of tester protein complexes and expressing the plurality of the target fusion proteins includes causing mating between first and second populations of haploid yeast cells of opposite mating types,

wherein

the first population of haploid yeast cells comprises

a library of tester expression vectors for the library of tester fusion proteins, each tester expression vector comprising

a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator,

a first nucleotide sequence encoding the first polypeptide subunit fused which is expression as a fusion protein with either the activation domain or the DNA binding domain of the transcription activator, and

a second nucleotide sequence encoding the second polypeptide subunit which is expressed as a separate protein from the first polypeptide subunit; and

the second population of haploid yeast cells comprises a plurality of target expression vectors, each of the target expression vector comprising

a second transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors, and

a target sequence encoding the target protein or peptide,

wherein either the first or second population of haploid yeast cells further comprises a reporter construct comprising the reporter gene whose expression is under transcriptional control of the transcription activator.

89. The method of claim 88, wherein members of the library of tester expression vectors are arrayed as individual yeast clones in one or more multiple-well plates.

90. The method of claim 87, wherein members of the library of target expression vectors are arrayed as individual yeast clones in one or more multiple-well plates.

91. The method of claim 87, wherein the mating is based on clonal mating in which each yeast clone containing members of the tester expression vectors is mated individually with each of the members of the library of target expression vector.

92. The method of claim 87, wherein the plurality of target expression vectors form a library of expression vectors containing a collection of human EST clones or a collection of domain structures.

93. A kit, comprising:

a first and second populations of haploid yeast cells of opposite mating types,

the first population of haploid yeast cells comprising

a library of tester expression vectors for the library of tester fusion proteins, each tester expression vector comprising

a first transcription sequence encoding either the activation domain or the DNA binding domain of the transcription activator,

a first nucleotide sequence encoding the first polypeptide subunit fused which is expression as a fusion protein with either the activation domain or the DNA binding domain of the transcription activator, and

a second nucleotide sequence encoding the second polypeptide subunit which is expressed as a separate protein from the first polypeptide subunit; and

the second population of haploid yeast cells comprising a target

expression vector, the target expression vector encoding

the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors, and

a target sequence encoding the target protein or peptide;

wherein either the first or second population of haploid yeast cells further comprising a reporter gene whose expression is under transcriptional control of the transcription activator.

94. The kit of claim 93, wherein the second population of haploid yeast cells comprises a plurality of target expression vectors, each of the target expression vectors encoding

the activation domain or the DNA binding domain of the transcription activator which is not expressed by the library of tester expression vectors; and

a target sequence encoding the target protein or peptide.

95. The kit of claim 93, wherein the haploid yeast cells of opposite mating types are $\underline{\alpha}$ and $\underline{a}$ type strains of yeast.

96. The kit of claim 93, wherein the first polypeptide subunit comprises an antibody heavy-chain region, and the second polypeptide subunit comprises an antibody light-chain region.